ESCAP II: Evaluation Results for Changes in A.C.E. Enumeration Status

Elizabeth A. Krejsa
and David A. Raglin
Planning, Research, and
Evaluation Division

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

EXECUTIVE SUMMARY

Are there errors in the identification of erroneous enumerations?

There are 1,919,029 net erroneous enumerations not found in production which were identified by the Measurement Error Reinterview (MER). The net erroneous enumerations is the difference between the correct enumerations that became erroneous in MER and the erroneous enumerations that became correct in MER. A majority of these additional erroneous enumerations were due to the identification of people who reported to have lived elsewhere in the Evaluation Followup Interview and geocoding errors not identified in production.

Background on the Evaluation Followup Interview (EFU) and the MER matching

- The EFU was conducted in January and February, 2001 via personal visit interviews in a 1-in-5 sample of the Accuracy and Coverage Evaluation clusters.
- The EFU collected information about people listed in either the census or the A.C.E. Person Interview, including inmovers.
- The EFU asked questions with the goal of determining if a person was a resident of the housing unit on census day according to the census residence rules.
- Using the EFU form, matchers determined the best residence and match code for a person. If the EFU form did not provide better information than was originally obtained in the Person Followup the residence status and match codes were not changed.

Comparison of Enumeration Status

- <u>Overall</u> - 3.5 percent of the EFU sample changed enumeration status between production and the evaluation.
- <u>Correct to Erroneous</u> – 2,827,414 production correct enumerations were coded erroneous enumerations in the MER.
- <u>Erroneous to Correct</u> – 908,385 production erroneous enumerations were coded correct enumerations in the MER.
- <u>Net Difference in Erroneous Enumeration Coding</u> – The net difference between the correct to erroneous and erroneous to correct enumerations is 1,919,029. This number is the additional erroneous enumerations found by the MER.
- <u>Unresolved Rate</u> – The unresolved rate following the MER matching was 1.7 percent. The unresolved rate for the production cases in the MER sample was 2.6 percent.

Source of changes in erroneous enumerations

Of the 2,827,414 people who were correctly enumerated according to production but erroneously enumerated according to the EFU:

- 92.2 percent were reported to have lived elsewhere,
- 5.0 percent were geocoding errors,

- 1.6 percent were duplicates,
- 0.9 percent were discrepant, and
- 0.3 percent were not able to be matched.

Analysis of demographic characteristics for persons reported to have lived elsewhere

At this time we cannot determine the specific information from the EFU form to indicate what type of place the person lived at when the person was coded as having lived elsewhere. However, we can explore demographics of the people who changed status.

- The poststratification age group with the highest percentage of correct enumeration to erroneous enumeration changes was the 18-29 year olds. While this group accounts for 14.8 percent of the population included in the MER, they, account for 34.1 percent of the changes. In addition, college-aged people (18-22 year olds) show a large concentration of coding errors, 19.3 percent of the correct enumerations to erroneous enumerations changes.

- The age group analysis was also reviewed for people coded as erroneous in production but changed to correct in the evaluation. People who are between 18-29 also account for a large proportion of the erroneous to correct enumerations changes (22.0 percent) relative to their presence in the population, though the differences between the age groups was not as large.

This suggests that the mobility of the 18-20 year old group may cause enumeration problems, and the effect on different enumeration statuses is similar.

Other demographic variables (such as sex, relationship, tenure, race, and poststrata) showed that nonrelatives of the household and minorities account for the largest percentage of the shift in enumeration status between correct and erroneous.

Correct enumeration probabilities for imputed unresolved cases

Unresolved production cases were sent for followup again in the EFU. One of the reasons for attempting to resolve these cases is to compare the results from the EFU to the correct enumeration probabilities production imputed. By conducting interviews again, the EFU was able to resolve 55.8 percent of the production unresolved cases. We determined that 75.6 percent of those resolved cases were correct enumerations. In production, all of these unresolved cases were imputed, resulting in 77.0 percent imputed as correct enumerations. Thus, the EFU results support the correct enumeration probabilities used in the production imputation process.

Review of the Results

Data for the MER, along with data from the Person Followup interview (PFU) were included in a second review, called the PFU and EFU Review, to assess the accuracy and validity of the MER and PFU data. The results of that review showed a decrease in the estimate of additional erroneous enumerations. See ESCAP II Report #24 "Results of the Person Followup and Evaluation Followup Forms Review", by Tamara Adams and Elizabeth A. Krejsa.

1. BACKGROUND

This report focuses on addressing the question: Are there errors in the identification of erroneous enumerations? We examine changes in enumeration status between the production matching process and an evaluation matching operation and define the sources of those changes.

1.1 Overview of the Production Operations

The Person Followup (PFU) interviewing phase of the Accuracy and Coverage Evaluation (A.C.E.) involved the followup of persons to resolve inconsistent information between the P-sample and the E-sample. The P-sample (Population Sample) consists of persons from the A.C.E. Person Interview who were listed on the A.C.E. Independent Roster and who were, according to census residence rules, members of the household on Census Day or whose residence status is unresolved. The P-sample is used to estimate missed people in the census. The E-sample (Enumeration Sample) comprises the final roster of persons from the census in the A.C.E. sample block clusters. The E-sample is used to estimate erroneous enumerations in the census.

After the A.C.E. housing unit and the person interviewing operations were completed, the person followup matching process was conducted. There were four major steps to the person followup matching process:

- Computer Match - The P-sample and the E-sample people were matched by computer. The results were used during the before followup clerical matching.

- Before Followup Matching–The clerical matchers reviewed the P-sample and E-sample persons whom the computer could not match, those who are possibly matched, and census cases with insufficient information for matching. The matchers also attempted to identify and code duplicated persons within both the P- and E-samples.

- PFU Interview - Unresolved and/or unmatched persons were selected for a field interview. During the interview additional information was obtained to help assign a final match and/or residency status to each person. For the E-sample, nonmatches were sent for a follow-up interview to determine if they were correctly or erroneously enumerated in the block cluster. Possible matches were also sent for an interview to resolve their match status.

- After Followup Matching –The information obtained in the PFU interview was used to code the match and/or residence or enumeration status of the persons in question.

1.2 Defining enumeration status

Match codes assigned to census people (in any of the matching steps) are grouped together into an enumeration status. The enumeration status of a person is classified as either correct, erroneous, or unresolved (Childers, 2000).

1.2.1 Correct enumerations

Correct enumerations (CEs) are people counted in the Census who are determined to have lived at the housing unit, which is determined to be in the search area, according to Census residence rules on Census Day. The Census person can be matched to an A.C.E. person or not matched.

1.2.2 Erroneous enumerations

Erroneous enumerations (EEs) are people counted in the Census who are determined to be:

- not matched and living elsewhere outside of the search area,
- living in a housing unit that was incorrectly classified as being in the search area,
- discrepant (i.e. possibly fictitious) in the search area,
- duplicated to another person,
- unmatchable because an incomplete name or invalid name was provided, or
- matched to an A.C.E. person but determined in a followup interview to be living elsewhere.

1.2.3 Unresolved cases

Unresolved people (URs) are those in which not enough information is collected during an interview to:

- identify the census person (either matched or not matched) as correctly or erroneously enumerated in the census,
- determine where the housing unit is located, or
- determine if a possible match is a match or not.

2. METHODS

This evaluation process focused on the interviewing and after followup matching steps in the person followup matching process.

2.1 The Evaluation Followup Interview

The data for this report were obtained from the Evaluation Followup Interview (EFU). The EFU is similar to the PFU in that it is a followup interview intended to resolve matching and residency issues.

Here are some facts about the EFU:

- The EFU was conducted in January and February, 2001 via personal visit interviews.
- EFU data were collected in a sample of about 1/5th of A.C.E. clusters in the country.
- The EFU collected information about people listed in either the census or the A.C.E. Person Interview. E-sample people sent to PFU in evaluation clusters were included in EFU as were a sample of matched people.
- The EFU asks questions about when people moved in and moved out of the sample housing unit as well as special living situations such as specific group quarters and other residences, with the goal of determining if a person was a resident of the housing unit on census day according to the census residence rules.

2.2 Measurement Error Reinterview matching

Using this expanded information from the EFU, the Measurement Error Reinterview (MER) matching process (structured similarly to the after followup matching step) was conducted to determine enumeration status of the persons in question.   The matchers could change the match code of a person from the production code based on the new information in the EFU.

3. LIMITS

The data in this report were obtained from the EFU. The most significant limitation of the EFU is the nine to ten month time lag between census day, April 1, 2000, and when the EFU data were collected in January and February, 2001. People move in that time period. People forget or inaccurately report information. The EFU questionnaire was developed, though, to attempt to minimize such problems by asking questions of the respondent that aid them in recalling the correct information. In addition, the clerical matchers had the option to reject the information on the EFU form and accept the production results if they did not believe the EFU data or the data were incomplete–this happened for roughly 10 percent of the people in EFU.

Another limitation is that the EFU did not have a full field quality assurance program as did the A.C.E. Person Interview and the PFU.

4. RESULTS

Only people who were sent for followup in EFU are included in the tables. All data provided are weighted to represent the EFU universe for the whole country. Standard errors are included in parentheses.   They are stratified jackknife estimates, based on the evaluation poststrata, and do not fully capture all phases of A.C.E. sampling.

4.1 Comparison of enumeration status

Table 1 below shows the change in enumeration status between production and EFU. These tables include all people who were followed up in EFU, including people for which the EFU form was rejected and the production code was retained.

Table 1. Enumeration Status for People who were Followed Up in EFU

	Evaluation			
Production	**Correct Enumerations**	**Erroneous Enumerations**	**Unresolved**	**Total**
Correct Enumerations	247,114,898 (6,337,607)	2,827,414 (223,469)	1,424,770 (254,488)	**251,367,081 (6,401,444)**
Erroneous Enumerations	908,385 (99,380)	3,118,191 (202,575)	124,641 (23,369)	**4,151,217 (239,619)**
Unresolved	2,873,110 (400,351)	928,719 (117,602)	3,010,280 (203,352)	**6,812,110 (491,207)**
Total	**250,896,393 (6,420,477)**	**6,874,324 (365,044)**	**4,559,691 (353,074)**	**262,330,408 (6,587,559)**

* Totals may not add due to rounding *

Overall, 3.5 percent of the weighted data changed enumeration status between production and the evaluation.   A sample of matches were sent to followup and are included in Table 1. In 1990, 13.1 percent of the weighted data changed enumeration status between production and the EFU. While a sample of matches were also sent to followup in 1990, they do not appear to be included in the E-sample analysis (West). Taking these matches out of the above table (matches remained correct enumerations in 98.8 percent of the cases), 15.0 percent of the weighted data changed enumeration status between production and EFU, a slightly higher rate than in 1990.

4.2 Correct enumeration probabilities for imputed unresolved cases

Unresolved production cases were sent for followup again in the EFU. One of the reasons for attempting to resolve these cases is to compare the results from the EFU to the correct enumeration probabilities production imputed.   By conducting interviews again, the EFU was able to resolve 55.8 percent of the production unresolved cases.   We determined that 75.6

percent of those resolved cases were correct enumerations. In production, all of these unresolved cases were imputed, resulting in 77.0 percent imputed as correct enumerations. Thus, the EFU results support the correct enumeration probabilities used in the production imputation process.

4.3 Changes in erroneous enumerations

As noted in section 1.2, there are various types of erroneous enumerations. Of the 3,118,191 erroneous enumerations that stayed erroneous enumerations, 96.9 percent were classified as the same type in production and in the evaluation.

4.3.1 Balancing erroneous enumeration errors

Of the changes affecting the erroneous enumerations, the largest is a change from correct enumerations. This change from CEs to EEs, however, is not balanced with changes from EEs to CEs. Since the change is not balanced, it appears as though production potentially over classified people as correct enumerations.

Why might these numbers be unbalanced? Specific information (such as answers to other residence questions or detailed notes as to where the person was living) is usually needed to code a person an EE. For example, if a respondent in production gave specific information indicating that a person was an EE and the EFU interviewer received less detailed information indicating that the person was a CE, the matcher may have decided to believe the production information. The matcher would then reject the EFU interview leaving the person an EE. So, it follows that persons determined to be EEs in production are less likely to be changed to CEs based on the EFU and that production CEs are more likely to change to EEs. To support this logic, we found that the EFU interview for cases that were determined to be EEs in production were rejected more often than the CEs, 17.5 percent of the time versus 11.2 percent. Therefore, it was determined that the PFU interview provided better information than the EFU.

4.3.2 Source of changes in erroneous enumerations

Knowing that this imbalance exists, it is useful to explore the sources of such changes in enumeration status. A change from CE to EE may happen if, for example, the detailed EFU prompted the respondent to remember that the followup person lived at college or had moved in after April 1st. We expected some of the correct enumerations in production to become erroneous enumerations in the EFU as a result of the inclusion of moving dates on the questionnaire, as well as detailed residency questions, and the addition of some households to a geocoding search that were not previously reviewed for such a purpose.

Based on match codes alone, of the 2,827,414 people who were correctly enumerated according to production but erroneously enumerated according to the EFU:

- 92.2 percent were reported to have lived elsewhere,

- 5.0 percent were geocoding errors,
- 1.6 percent were duplicates,
- 0.9 percent were discrepant, and
- 0.3 percent were not able to be matched.

4.3.3 Analysis of demographic characteristics for persons reported to have lived elsewhere

At this time we cannot determine the specific information from the EFU form to indicate what type of place the person lived at when the person was coded as having lived elsewhere. However, we can explore demographics of the people who changed status.

The poststratification age group with the highest percentage of CE to EE changes was the 18-29 year olds. (Note that we also explored the percent for those coded as 'living elsewhere' for each age group and found the same results as above.) It is likely that this difference is because either they were college students (living in a dorm or other housing) or because this group is highly mobile relative to other portions of the population. Separating these differences at this time is not possible. Based on age alone, college-aged people (18-22 year olds) show the largest concentration of coding errors. As seen in Table 2 below, 18-29 year olds are the largest portion of the CEs to EEs cell, with 18-22 year olds accounting for 19.3 percent.

Table 2. Percent of Change by Age Group

Age Group	% of Total Population (N=262,330,408)	% of Production CEs that became EEs in EFU (N=2,827,414)	% of Production EEs that became CEs in EFU (N=908,385)
0-17	26.2	17.0 (2.1)	23.4 (3.0)
18-29	14.8	34.1 (3.6)	22.0 (2.9)
30-49	31.0	18.9 (2.4)	25.8 (2.4)
50+	28.0	30.0 (3.5)	28.8 (3.5)
Total	100.0	100.0	100.0

The age group analysis was also reviewed for people coded as erroneous in production but changed to correct in the evaluation. People who are between 18-29 again account for a large proportion of the cell relative to their presence in the population, though the differences between groups was not as large. In general, this suggests that the mobility of this group causes enumeration problems, and it appears to affect different enumeration statuses similarly.

Table 3 shows the percent of change by sex. The table indicates that men were relatively more likely to change enumeration status than women.

Table 3. Percent of Change by Sex

Sex	% of Total Population (N=262,330,408)	% of Production CEs that became EEs in EFU (N=2,827,414)	% of Production EEs that became CEs in EFU (N=908,385)
Male	48.4	50.6 (2.9)	54.3 (2.3)
Female	51.6	49.4 (2.9)	45.7 (2.3)
Total	100.0	100.0	100.0

The analysis for relationship indicates that other relatives and nonrelatives were relatively more likely to have a change in enumeration status, which is not a surprise. These people may have less of an attachment to the household than the reference person, a spouse, or children of the reference person. They also may be more mobile than other household members.

Table 4. Percent of Change by Relationship

Relationship	% of Total Population (N=262,330,408)	% of Production CEs that became EEs in EFU (N=2,827,414)	% of Production EEs that became CEs in EFU (N=908,385)
Spouse	20.3	12.8 (2.2)	14.2 (1.9)
Child	30.7	35.7 (3.2)	32.7 (2.8)
Sibling	1.1	3.3 (1.6)	1.9 (0.8)
Parent	1.2	1.7 (0.5)	2.6 (1.2)
Other relative	3.1	4.7 (0.9)	6.1 (1.6)
Nonrelative	4.8	10.3 (2.2)	7.9 (1.6)
Reference Person	38.8	31.5 (2.8)	34.6 (1.5)
Total	100.0	100.0	100.0

Table 5 shows that renters were more likely to change residence status. Renters are more mobile and less attached than owners, which can make them more difficult to count at the right location.

Table 5. Percent of Change by Tenure

Tenure	% of Total Population (N=262,330,408)	% of Production CEs that became EEs in EFU (N=2,827,414)	% of Production EEs that became CEs in EFU (N=908,385)
Owner	71.4	63.0 (3.9)	56.5 (5.8)
Renter	28.6	37.0 (3.9)	43.5 (5.8)
Total	100.0	100.0	100.0

Table 6 examines race. It shows that whites were less likely to change enumeration status and people who were not white or other were more likely to change enumeration status.

Table 6.   Percent of Change by Race

Race	% of Total Population (N=262,330,408)	% of Production CEs that became EEs in EFU (N=2,827,414)	% of Production EEs that became CEs in EFU (N=908,385)
White	77.4	70.8 (3.3)	69.7 (4.3)
Other race OR White and other	5.6	5.1 (0.8)	7.8 (2.0)
Remaining	17.0	24.1 (3.2)	22.5 (3.6)
Total	100.0	100.0	100.0

4.4   Changes in geocoding

Geocoding of housing units has been a concern. To address this concern, additional Targeted Extended Search (TES) work was conducted during the EFU interview. This work was redone for some housing units previously targeted in the production TES as well as for housing units not previously included. In addition to conducting a search of the surrounding rings, some housing units were reviewed simply to determine if they were inside or outside of the cluster.

In cases where some type of geocoding work was conducted, 28 percent (unweighted) of the EFU cases had a geocoding conflict with the production geocoding results. In those cases we decided after research that the EFU geocoding was correct. In addition, there were many times when EFU and production information disagreed but matchers believed that the production information was correct and rejected the EFU results. Unfortunately, we are unable to determine how often this happened. The results of this work show that, in general, interviewers (regardless of the operation) have a difficult time locating housing units.

4.4.1 Redone TES work

Table 7. Comparison of TES

	Evaluation	
Production	Non-Geocoding	Geocoding
Non-Geocoding	5760	95
Geocoding	39	563

When both production and the evaluation conducted TES, 19 percent conflicted (39+95/563+39+95) with adding of geocoding errors occurring more often.

4.4.2 EFU TES cases that were previously not searched for in the surrounding rings

Table 8. Comparison of EFU TES and Production Geocoding

	Evaluation	
Production	Non-Geocoding	Geocoding
Non-Geocoding	1630	39
Geocoding	88	79

Note that the 88 cases that were geocoding errors in production were not searched for in the surrounding ring. Production classified these cases as unresolved geocoding problems and the correct enumeration probability was imputed. Therefore, the work done in production noted that the housing unit was not in the cluster but could not determine how far outside of the cluster it was. The EFU confirmed that the housing unit was not in the cluster and determined that it was in the first ring surrounding the cluster.   This supports the decision made by production to not treat these cases as geocoding errors.

4.4.3 Redone Geocoding work

Table 9.   Comparison of Geocoding

	Evaluation	
Production	Non-Geocoding	Geocoding
Non-Geocoding	2156	124
Geocoding	28	405

When both production and the evaluation conducted a geocoding check, 27 percent conflicted with adding of geocoding errors occurring more often. Note that these are not TES clusters. The interviewers were strictly determining if the housing unit is in the cluster or outside of the cluster.

5. CONCLUSIONS

Results of this study showed a net shift of approximately 1.9 million EEs to CEs from production. The number of CEs has a direct effect on the dual system estimate (DSE). If this is an accurate measure, then the production DSE would be too high.

This increase in EEs seemed large so additional review of the forms was deemed appropriate. A sample of the Person Followup (PFU) and EFU forms for the E-sample were reviewed to ascertain the correct code based on each form and the reason for assigning that code.   The results of this review were documented by Adams and Krejsa (2001).

6. REFERENCES

Adams, Tamara, and Elizabeth Krejsa (2001). "ESCAP II: Results of the Person Followup and Evaluation Followup Forms Review", ESCAP II Report #24.

Childers, Danny R (2000). "Accuracy and Coverage Evaluation: The Design Document." DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-01. October 11, 2000.

West, Kirsten. "1990 Post-Enumeration Survey Evaluation Proejct P9a: Accurate Measurement of Census Erroneous Enumerations -- Evaluation Followup." 1990 Coverage Studies and Evaluation Memorandum Series # K-2. July 11,1991.